Evaluation Document

Matan Ohayon & Itay Nov  
ID : 207696907 , 208327163

1. The classes we will examine will be classified based on "Machine failure," with a division of 0 or 1.  
   Findings:  
   Number of findings classified as '0': 134,281  
   Number of findings classified as '1': 2,148
2. The regression equation (according to section 7):

**Coefficients**:

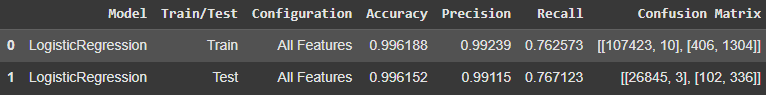
*[[-0.42724932 2.6515951 -1.54361874 1.91234209 5.99637625 1.25467066*

*8.80666 8.08065176 9.17788897 7.74249689 -0.01684327]]*

**Intercept**:

*-9.936131198261497*

**Accuracy Metrics and Confusion Matrix**:



1. The regression equation (according to section 8):

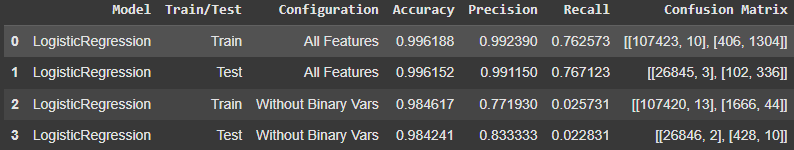
**Coefficients**:

*[[-0.21341729 5.65686931 -4.23340376 4.87281891 11.28798545 1.94640648]]*

**Intercept**:

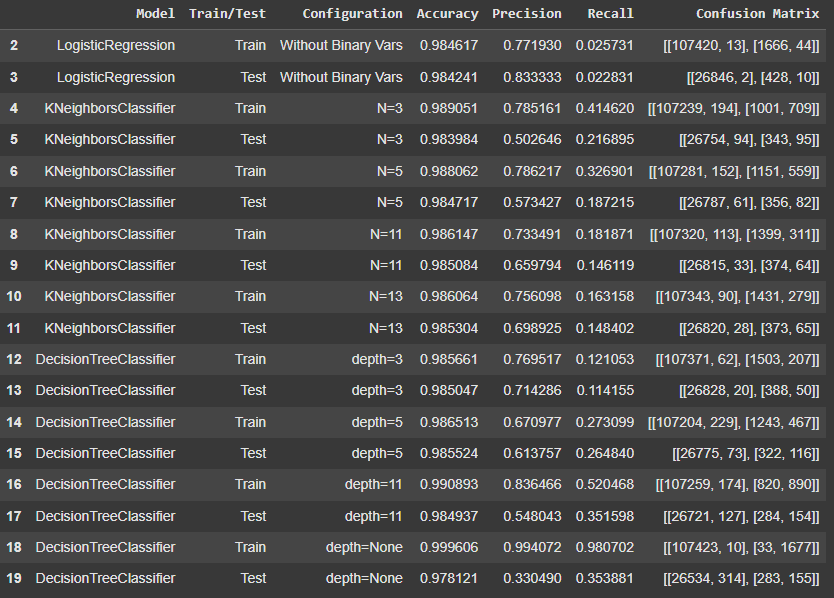
*-12.978196185960767*

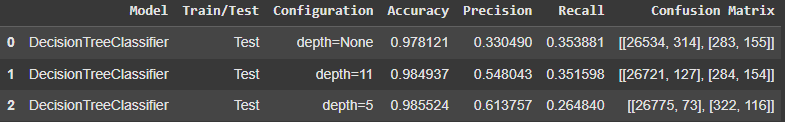
**Accuracy Metrics and Confusion Matrix**:



Impact: Removing these features significantly reduced the model’s ability to identify failures (recall dropped from 0.767 to 0.022). This demonstrates the critical importance of these features for the prediction task.

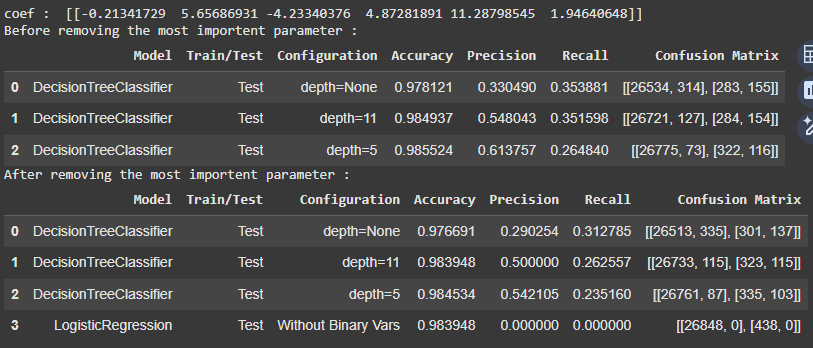
1. The additional models tested in section 9 are: KNN and Decision Tree.  
   Each model was tested with 4 different configurations:  
   a. **KNN**: The data was tested for 3, 5, 11, and 13 neighbors.  
   b. **Decision Tree**: The data was tested for different depths: 3, 5, 11, and with no depth limit
2. The accuracy metrics for the train and test sets of the regression model from section 8, and the models from section 9 (including the hyperparameter values in their configurations), are presented in the following table :



1. The Top preforming models :

We chose these models because they provided the highest Recall, which is important for ensuring that failures are detected. A high Recall minimizes the risk of missing critical failure events, which is crucial in predicting machine breakdowns.

1. The variable with the highest importance in the regression equation is 'Torque [Nm]'. Below are the model metrics before and after removing this variable:



Observation:

1. Accuracy: Minimal impact, remaining high (~0.97–0.98).
2. Precision: Drops across all Decision Tree configurations, most notably at depth=None. Logistic Regression fails entirely (Precision = 0).
3. Recall: Decreases across the board, indicating the parameter's critical role in identifying positives.
4. Logistic Regression: Heavily reliant on the removed parameter, failing to detect positives without it.
5. Decision Trees: More robust, retaining high accuracy, but with reduced precision and recall.